

SEMICONDUCTOR DEVICE

Patent number:	JP5335656
Publication date:	1993-12-17
Inventor:	SHIBATA SUNAO; others: 04
Applicant:	SUNAO SHIBATA; others: 02
Classification:	
- international:	H01L49/00; H01L27/10
- european:	
Application number:	JP19920160187 19920527
Priority number(s):	

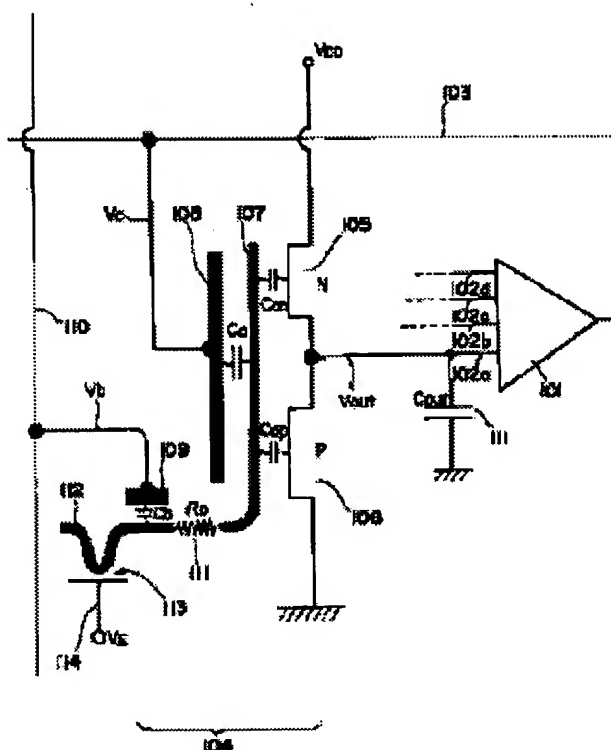
Also published as:

 WO9324957 (A1)

Abstract of JP5335656

PURPOSE: To achieve synapse connection with a low power consumption and a small number of elements by allowing a resistance at a node which is connected to the extended part of a second floating gate electrode to have a value which is larger than the operation resistance value of a tunnel connection which feeds charge to and obtains it from a third floating gate electrode part.

CONSTITUTION: A high-resistance node 111 is placed between a MOS-type transistor floating gate 107 and a second floating gate 112 and a constant amount of charge is injected to the second floating gate 112 for each pulse which is applied to an input gate 109. The charge is slowly fed to the floating gate 107 of a synapse circuit 104 via the high-resistance node 111. As a result, the constant amount of charge can be injected for each pulse. The injection of charge to the floating gate or the release of charge of a floating gate 103 can be accurately controlled according to the number of pulses and the weight value of synapse can be controlled precisely.



Data supplied from the **esp@cenet** database - Patent Abstracts of Japan

(51)Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
H 0 1 L 49/00		Z 8728-4M		
27/10	4 5 1	8728-4M		

審査請求 未請求 請求項の数15(全 21 頁)

(21)出願番号 特願平4-160187

(22)出願日 平成4年(1992)5月27日

(71)出願人 591022117

柴田 直

宮城県仙台市太白区日本平5番2号

(71)出願人 000205041

大見 忠弘

宮城県仙台市青葉区米ヶ袋2-1-17-301

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72)発明者 柴田 直

宮城県仙台市太白区日本平5番2号

(74)代理人 弁理士 福森 久夫

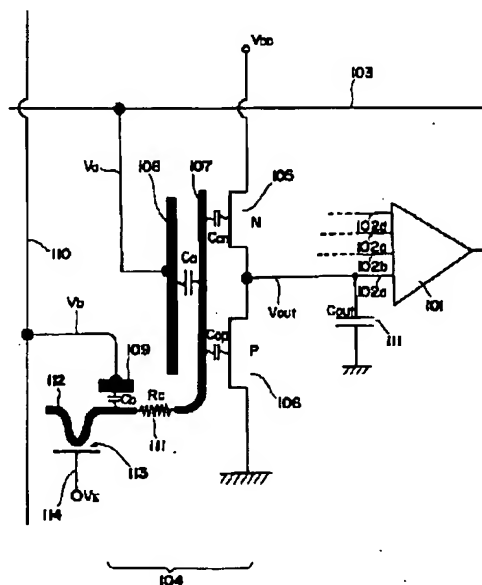
最終頁に続く

(54)【発明の名称】 半導体装置

(57)【要約】 (修正有)

【目的】少数の素子によってシナプス結合が構成でき、電力消費が非常に少なく、神経回路網の高集積化、低電力化が可能となる。高精度のシナプス加重値の変更ができ、実用的なレベルのニューロンコンピュータチップを実現できる。

【構成】 フローティングゲート電極と第2絶縁膜を介して容量結合する複数の第1の入力ゲート電極を有し、前記第1の入力ゲート電極の1つにソース電極が接続された第1のMOS型トランジスタを有し、第1のMOSには電位的にフローティング状態にある第2のフローティングゲート電極が設けられ、第2のフローティングゲート電極の延在部に接続部を介して電氣的に接続された第3のフローティングゲート電極を有し、第3のフローティングゲート電極部に電荷を出し入れするトンネル接合部を有するMOS型半導体装置において、少なくとも接続部の抵抗が前記トンネル接合の動作抵抗値よりも大きな値を有するよう構成する。



【特許請求の範囲】

【請求項1】 基体上に一導電型の第1の半導体領域を有し、この領域内に設けられた反対導電型の第1のソース及び第1のドレイン領域を有し、前記第1のソース、及び第1のドレイン領域を隔てる領域に第1の絶縁膜を介して設けられた電位的にフローティング状態にある第1のフローティングゲート電極を有し、前記第1のフローティングゲート電極と第2の絶縁膜を介して容量結合する複数の第1の入力ゲート電極を有し、前記第1の入力ゲート電極の1つにソース電極が接続された第1のMOS型トランジスタを有し、前記第1のMOS型トランジスタには電位的にフローティング状態にある第2のフローティングゲート電極が設けられ、前記第2のフローティングゲート電極の延在部に接続部を介して電気的に接続された第3のフローティングゲート電極を有し、前記第3のフローティングゲート電極部に電荷を出し入れするトンネル接合部を有するMOS型半導体装置において、少なくとも前記接続部の抵抗が前記トンネル接合の動作抵抗値よりも大きな値を有するよう構成されたことを特徴とする半導体装置。

【請求項2】 前記接続部が、前記第2及び／もしくは前記第3のフローティングゲート電極とは異った抵抗率をもった材料で構成されていることを特徴とする請求項1に記載の半導体装置。

【請求項3】 前記接続部、前記第2及び第3のフローティングゲート電極が略々等しい抵抗率をもった材料で構成されていることを特徴とする請求項1に記載の半導体装置。

【請求項4】 前記接続部に第3の絶縁膜を介して設けられたゲート電極を有し、前記ゲート電極に加えられる電圧により、前記接続部の抵抗値が変化するよう構成されたことを特徴とする請求項1乃至請求項3のいずれか1項に記載の半導体装置。

【請求項5】 前記第1のMOS型トランジスタが、前記第2のフローティングゲート電極と第4の絶縁膜を介して容量結合する少なくとも一個の第2の入力ゲートを有し、前記第2の入力ゲートに高レベル又は低レベルの2つの電位レベルの信号を伝達する第1の配線が接続されていることを特徴とする請求項1乃至請求項4のいずれか1項に記載の半導体装置。

【請求項6】 前記第1のMOS型トランジスタがNチャネル型であり、そのソース電極がPチャネル型の第2のMOS型トランジスタを介して、低レベルの電圧を供給する電源ラインに接続されていることを特徴とする請求項1乃至請求項4のいずれか1項に記載の半導体装置。

【請求項7】 前記第1のMOS型トランジスタがPチャネル型であり、そのソース電極がNチャネル型の第3のMOS型トランジスタを介して、高レベルの電圧を供給する電源ラインに接続されていることを特徴とする請

求項1乃至請求項6のいずれか1項に記載の半導体装置。

【請求項8】 前記第2のMOS型トランジスタが第4のフローティングゲート電極を有し、前記第2及び第4のフローティングゲート電極が電気的に接続されていることを特徴とする請求項6に記載の半導体装置。

【請求項9】 前記第1のMOS型トランジスタのソース及びドレインのいずれか一方の電極がNチャネル型の第4のMOS型トランジスタを介して低レベルの電圧を供給する電源ラインに接続され、もう一方の電極がPチャネル型の第5のMOS型トランジスタを介して高レベルの電圧を供給する電源ラインに接続されていることを特徴とする請求項1乃至請求項5のいずれか1項に記載の半導体装置。

【請求項10】 前記第4及び第5のMOS型トランジスタのゲート電極が互に電気的に接続されていることを特徴とする請求項9に記載の半導体装置。

【請求項11】 前記第1のMOS型トランジスタのソース及び基体が互に電気的に接続されていることを特徴とする請求項1乃至請求項10のいずれか1項に記載の半導体装置。

【請求項12】 前記第1及び第2のMOS型トランジスタの半導体基体がそれぞれのソース電極と電気的に接続されていることを特徴とする請求項6乃至請求項8のいずれか1項に記載の半導体装置。

【請求項13】 前記第1の配線が互に平行に複数本配置され、それと直交する方向に前記第1のフローティングゲート電極が互に平行に複数本配置されたことを特徴とする請求項1乃至請求項12のいずれか1項に記載の半導体装置。

【請求項14】 前記第1のフローティングゲート電極が少なくとも1つのMOS型トランジスタのソースもしくはドレインに接続されていることを特徴とする請求項1乃至請求項13のいずれか1項に記載の半導体装置。

【請求項15】 前記第2及び第4の絶縁膜が、前記第1の絶縁膜の誘電率よりも大きな誘電率を有する絶縁材料で形成されていることを特徴とする請求項1乃至請求項14のいずれか1項に記載の半導体装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、半導体装置に係わり、特に神経回路網コンピュータ（ニューロンコンピュータ）を実現するための高機能半導体集積回路装置を提供するものである。

【0002】

【関連技術】半導体の集積回路技術は実に驚くべき速度で進んでおり、例えばダイナミック・メモリを例にとるなら、4メガビットから16メガビットがすでに量産体制にあり、64メガビット以上の容量をもった超々高密度メモリも研究レベルでは実現されつつある。64メガ

ビットメモリは、せいぜい1cm四方のシリコンチップ上に実に約1億2000万個のMOSトランジスタが集積されている。このような超高集積化技術はメモリ回路ばかりでなく論理回路にも応用され、32ビットから64ビットのCPUをはじめとする、様々な高機能論理集積回路が開発されている。

【0003】しかし、これらの論理回路はデジタル信号、すなわち「1」と「0」という2値の信号を用いて演算を行なう方式を採用しており、例えばコンピュータを構成する場合は、ノイマン方式といって、あらかじめ決められたプログラムに従って1つ1つ命令を実行していく方式である。このような方式では単純な数値計算に対しては非常に高速な演算が可能であるが、パタン認識や画像処理といった演算には膨大な時間を要する。さらに、連想、学習といったいわば人間が最も得意とする情報処理に対しては非常に不得手であり、現在様々なソフトウェア技術の研究が行なわれているが、はかばかしい成果は得られていないのが現状である。ここで、これらの困難を一挙に解決するため、生物の脳の機能を研究し、その機能を模倣した演算処理の行なえるコンピュータ、すなわち神経回路コンピュータ（ニューロンコンピュータ）を開発しようというまた別の流れの研究がある。

【0004】このような研究は、1940年代より始まっているが、ここ数年来非常に活発に研究が展開されるようになった。それはLSI技術の進歩にともない、このようなニューロンコンピュータのハードウェア化が可能となったことによる。

【0005】しかしながら、現状の半導体LSI技術を用いてニューロンコンピュータをLSIチップ化するにはまだまだ様々な問題があり、実用化のメドはほとんどたっていないのが実情である。

【0006】LSI化における技術的な問題がどこにあるのかを以下に説明する。

【0007】人間の脳は極めて複雑な構造を有し、非常に高度な機能を有しているが、その基本的な構成は非常に単純である。すなわち、ニューロンと呼ばれる演算機能をもった神経細胞と、その演算結果を他のニューロンに伝える、いわば配線の役割を担った神経繊維とから構成されている。

【0008】この脳の基本単位の構成を簡略化してモデルで描いたのが図9である。901a, 901b, 901cはニューロンであり、902a, 902b, 902cは神経繊維である。903a, 903b, 903cはシナプス結合とよばれ、例えば神経繊維902aを伝わって来た信号に w_a という重みをかけ、ニューロン901aに入力する。ニューロン901aは入力された信号強度の線形和をとり、それらの合計値がある閾値をこえると神経細胞が活性化し、神経繊維902bに信号を出力する。合計値が閾値以下だとニューロンは信号を出力

しない。合計値が閾値以上になって、ニューロンが信号を出すことを、そのニューロンが「発火した」と言う。

【0009】実際の脳では、これらの演算、信号の伝搬、重みのかけ算等すべて電気化学現象によって行われており、信号は電気信号として伝送・処理されている。人間が学習する過程は、シナプス結合における重みが増加していく過程としてとらえられている。すなわち、様々な入力信号の組合せに対し、正しい出力が得られるよう重みが徐々に修正され、最終的に最適の値に落ち着くのである。つまり人間の英知はシナプスの重みとして脳に刻みつけられているのである。

【0010】数多くのニューロンがシナプスを介して相互に接続され1つの層を形成している。これらが人間の脳では、6層重ね合わされていることが分かっている。このような構造、機能を半導体デバイスを用いてLSIシステムとして実現することが、ニューロンコンピュータ実現の最も重要な課題である。

【0011】図10(a)は、1つの神経細胞、すなわち1個のニューロンの機能を説明する図面であり、1943年にMcCullockとPitts(Bull:Math. Biophys. Vol. 5, p.115(1943))により数学的モデルとして提案されたものである。現在もこのモデルを半導体回路で実現し、ニューロンコンピュータを構成する研究が盛んに進められている。 $V_1, V_2, V_3, \dots, V_n$ は、例えば電圧の大きさとして定義される n 個の入力信号であり、他のニューロンから伝達された信号に相当している。 $w_1, w_2, w_3, \dots, w_n$ はニューロン同士の結合の強さを表す係数で、生物学的にはシナプス結合と呼ばれるものである。ニューロンの機能は各入力 V_i に重み w_i ($i=1 \sim n$)をかけて線形加算した値 Z が、ある所定の閾値 V_{TH}^* より大となったときに「1」を出力し、また閾値より小のときに「0」を出力するという動作である。

これを数式で表せば、

【0012】

【数1】

$$Z \equiv \sum_{i=1}^n w_i V_i \quad \dots (1)$$

$$V_{out} = 1 \quad (Z > V_{TH}^*) \quad \dots (2)$$

$$0 \quad (Z < V_{TH}^*) \quad \dots (3)$$

となる。

【0013】図10(b)は、 Z と V_{out} の関係を表したものであり、 Z が V_{TH}^* より十分大きいときは1、十分小さいときは0を出力している。

【0014】さて、このようなニューロンをトランジスタの組合せで実現しようと思えば、数多くのトランジスタを必要とするばかりか、加算演算を各信号を電流値に変換してこれを足し合わせるにより行うため、多くの電流が流れ多大のパワーを消費することになる。これでは高集積化は不可能である。この問題は、ニューロン

MOSFET (νMOSと略) の発明 (発明者: 柴田直、大見忠弘、特願平1-141463号) により解決された。

【0015】この発明はたった1つのトランジスタでニューロンの働きの主要機能を果たすことができ、しかも電圧信号をそのまま加算演算することができるため、ほとんど電力消費がないという画期的なものである。図11(a)はνMOS断面構造の一例を簡略化して示したものであり、1101は例えばP型のシリコン基板、1102, 1103はN⁺拡散層で形成されたソース及びドレイン、1104はチャネル領域上に設けられたゲート絶縁膜 (例えばSiO₂など)、1106は電氣的に絶縁され電位的にフローティングの状態にあるフローティングゲート、1107は例えばSiO₂等の絶縁膜、1108 (G₁, G₂, G₃, G₄) は入力ゲートでありニューロンの入力に相当する。

【0016】図11(b)はその動作を説明するためにさらに簡略化した図面である。各入力ゲートとフローティングゲート間の容量結合係数をC_G、フローティングゲートとシリコン基板間の容量結合係数をC_Oとすると、フローティングゲートの電位Zは、

$$Z = -w(V_1 + V_2 + V_3 + V_4) \quad \dots (4)$$

$$W \equiv C_G / (C_O + 4C_G) \quad \dots (5)$$

とあらわされる。但しここで、V₁, V₂, V₃, V₄はそれぞれ入力ゲートG₁, G₂, G₃, G₄に入力されている電圧であり、シリコン基板の電位は0V、すなわちアースされているとした。

【0017】このνMOSはフローティングゲートをゲート電極とみれば通常のNチャネルMOSトランジスタであり、このフローティングゲートからみた閾電圧 (基板表面に反転層が形成される電圧) をV_{TH}^{*}とすると、Z > V_{TH}^{*}で上記νMOSはオンし、Z < V_{TH}^{*}ではオフする。つまりこのνMOS 1109を1つ用いて例えば同図(c)のようなインバータ回路を組めば簡単に1ヶのニューロンの機能が表現できるのである。1110, 1111はインバータを構成するための抵抗、1112はNMOSトランジスタである。同図(d)は、V_{out1}, V_{out2}をZの関数として示したものであり、Z > V_{TH}^{*}の入力に対しV_{out2}はV_{DD}のハイレベルの電圧を出力している。つまりニューロンが発火した状態を実現している。

【0018】(4)式で示したように、ニューロンへの入力が電圧レベルで加算され、その線形和が閾値以上になるとニューロンが発火するという基本的な動作がたった1つのνMOSによって実現されているのである。電圧モードの加算を行なうので、入力部で流れる電流はコンデンサの充放電電流のみであり、その大きさは非常に小さい。一方、インバータでは、ニューロン発火時に直流電流が流れるが、これは、負荷として、抵抗1110を用いているためであり、前記発明 (特願平1-141

463号) によるCMOS構成のνMOSゲートを用いれば、この直流電流はなくすることができる。

【0019】図12は、CMOS構成の一例を示す図面である。図12(a)はCMOSニューロンゲートの断面構造を模式的に表したものであり、1201はP型シリコン基板、1202はn型のウェル、1203a, 1203bはそれぞれN⁺型のソース及びドレイン、1204a, 1204bはそれぞれP⁺型のソース及びドレイン、1205はフローティングゲート、1206a~dはそれぞれ入力ゲートの電極である。1207, 1208は例えばSiO₂等の絶縁膜、1209はフィールド酸化膜である。同図(b)は1個のニューロン回路を構成した例であり、1210は同図(a)のCMOSニューロンゲートを記号であらわしたものであり、符号を付した部分は同図(a)の番号と対応している。1211はCMOSのインバータであり、1212, 1213はそれぞれNMOS及びPMOSのトランジスタである。また、1214はニューロンの出力である。

【0020】以上の様に、少数の素子で1ヶのニューロンが構成でき、しかもパワー消費が非常に少ないためνMOSはニューロンコンピュータを実現する上で、不可欠な素子となっているのである。

【0021】しかしながら、ニューロンコンピュータを実現するには、ニューロン以外のもう1つ重要な要素、すなわちシナプスも構成する必要がある。図13は、従来技術で構成したシナプス結合も含むニューロン回路の基本構成の一例である。

【0022】1301は例えば図11(c)に示したようなニューロン回路であり、1302は他のニューロンの出力信号を伝える配線である。1303はシナプス結合回路であり、入力信号に重みを付与するための回路である。NMOSトランジスタ1304のソース1306に負荷抵抗 (R+R_x) を接続したソースフォロワー回路となっている。従って、NMOSトランジスタのゲート電極1305に発火したニューロンの出力電圧V_sが印加されると、ソース1306には、V_s-V_{TH}なる電圧が出てくる (ここでV_{TH}は、NMOSトランジスタ1304の閾電圧である。)。

【0023】例えば、V_{TH}=0のMOSトランジスタを用いたとすると、ソース1306の電位はV_sと等しくなり、この電圧が2つの抵抗R, R_xで分割されてシナプス結合回路の出力電圧となり、結線1307によってニューロン1301に伝えられる。この出力電圧は、V_s・R_x / (R+R_x) となり、R_x / (R+R_x) なる重みが信号電圧V_sに掛けられたことになる。R_xの値を変換することにより重みを変更することができる。

【0024】図14(a)は可変抵抗の実現方法の一例を示したものである。例えば、1つのMOSトランジスタ1401のゲートに一定の電圧V_{GG}を印加してやれ

ば、このトランジスタは1つの抵抗の働きをする。 V_{GG} の値を変化させることによりその抵抗値を変化させることができる。

【0025】また、同図(b)は、 V_{GG} の値を制御する回路の一例を示したもので、4ビットのバイナリカウンタ1402とD/Aコンバータ1403とから構成されている。シナプスの結合強さは、4ビットの2進数で表現され、それが、D/Aコンバータ1403によってアナログ電圧に変換され V_{GG} の値として出力される。シナプス結合強度を強めるには、制御信号によりカウンタの値をカウントダウンさせ、 V_{GG} の値を小さくすればよい。逆にシナプス結合強度を弱めるには、カウントアップさせ、 V_{GG} の値を大きくしてやればよい。

【0026】さて、図13及び図14に示したようなシナプス結合回路を用いた場合の問題点を次に説明する。

【0027】まず、第1の問題点は図13で重みを発生させるのに抵抗による電圧分割を用いている点である。この方式ではこの抵抗に常に電流を流し続けることによって、重みを掛けた出力電圧を保持しているため、常時 $V_s^2 / (R + R_x)$ の電力を消費することになる。これでは、たとえニューロン1301における消費電力を $vMOS$ の応用により減少させても回路全体としての消費電力は決して小さくならない。一層が n 個のニューロンからなる2層の神経回路網を考えると、シナプス結合の数は n^2 個となりニューロンの数よりシナプスの数の方が圧倒的に多いのである。従って、常時電流を流し続けなければならないシナプス結合回路を用いる限り、実用的な規模の神経回路網を構成することは消費電力が過大となり、事実上設計不可能となっている。 $R + R_x$ の値を十分大きくすることにより消費電力を減少させることはできるが、こうした場合 C_{out} を充放電するための時定数が非常に大きくなり、シナプス回路の動作速度が著しく劣化することになる。

【0028】第2の問題点は、結合の問題の重みを決める、図14(b)に示した回路が多数の素子を必要とし、高集化できないという事実である。学習機能を有する神経回路網を構成するためには、各シナプス結合の強さは適宜変更ができ、かつその変更した値を記憶しておく必要がある。同図では、このために4ビットのバイナリカウンタを用いているが、これだけでも最低30個程度の MOS トランジスタを必要とする。さらにD/Aコンバータを構成するためにも多くの素子を必要とする。さらにこれらの回路が、1つのシナプス結合当たり、さらに多くの電力を消費することになり、消費電力の面からも不利となるのである。

【0029】シナプス構成に必要な素子数を低減させる方法として、フローティングゲート型の $EPROM$ や $E^2 PROM$ の不揮発性メモリを用いる方法が提案されている。これらのデバイスは、フローティングゲート内の電荷の量によって、その閾値が変化するため、電荷の量

によってアナログ的に重みを記憶することができる。一個のトランジスタで重みを記憶できるため1つ1つのシナプス回路は、図14(b)の回路にくらべて小さくすることができる。しかしながら、これを重みとして読み出し、前段のニューロンの出力に乗算するためには、やはり相当複雑な回路を必要とする。例えば、2つの $E^2 PROM$ メモリセルを用いた差動増幅回路を構成し[D. Soo and R. Meyer, "A Four-Quadrant NMOS Analogue Multiplier," IEEE J. Solid State Circuits, Vol. sc-17, No. 6, Dec., 1982]、重みを掛けた結果を電流信号として読み出すことになる。回路の大幅な単純化を達成できないばかりか、常時電流を流すことにより重みの掛け算を行うため消費電力が非常に大きくなり、やはり大規模ニューラルネットワーク構成には用いることができない。

【0030】さらに重大な問題点を図15に示す。

【0031】図15(a)はトンネル接合を有する $E^2 PROM$ セルの閾電圧(V_{TH})を、データ書き込み用のパルスの数の関数として示したものである。プログラム電圧は19.5Vであり、パルスの幅、5msecである。プログラム用の制御電極に正のパルスを加えると電子がフローティングゲート内に注入され閾値は正方向にシフトする。逆に負のパルスを印加すると電子がフローティングゲートから放出されて、閾値は負の方向にシフトする。図から明らかな様に、最初の一個のパルスによって閾値は大きくシフトし、その後のパルスによっては非常にわずかにしか変化していないことが分る。これでは、閾値を細く変化させて、シナプスの重みを、数多くのレベルに調整することは不可能である。

【0032】この原因は次の様に説明することができる。

【0033】図15(b)は、正のプログラム電圧をステップ関数的に印加したときの、フローティングゲートに注入される電子の数(n)の時間変化の様子を示したものである。電圧印加の初期に数多くの電子が注入され、その後はほとんど増加しないことが分る。これは、電荷注入の基本となっている。絶縁膜中を流れるFowler-Nordheim Tunnelingという電流が、

【0034】

【数2】

$$I \propto V^2 \exp\left(-\frac{b}{V}\right)$$
という式に従って、絶縁膜両端の電位差 V に依存するためである。即ち、初期のトンネル電流によってフローティングゲート内の電子の数が増加すると、これによってフローティングゲートの電位が下り、 V が小さくなり、その結果としてトンネル電流が指数関数的に減少してしまうからである。トンネル電流を一定値に制御し、シナプス加重を精度よく変更するには、フローティングゲート内の電荷の数に応じてパルス電圧の大きさやパルス幅を精度よくコントロールする必要があり、さらに多くの

回路を要する結果となる。

【0035】要するに、従来知られた技術では、低消費電力化、高集積化、さらにシナプス加重の精度のいずれの面からも神経回路網の構成はほとんど不可能と言わざるを得ない。従って、従来の技術ではニューロンコンピュータを実現することはできないのである。

【0036】

【発明が解決しようとする課題】そこで本発明は、このような問題点を解決するためになされたものであり、消費電力が非常に小さく、かつ少数の素子でシナプス結合が実現でき、高集積度、シナプス加重の高精度、低消費電力のニューロンコンピュータチップを実現することのできる半導体装置を提供するものである。

【0037】

【課題を解決するための手段】本発明の半導体装置は、基体上に一導電型の第1の半導体領域を有し、この領域内に設けられた反対導電型の第1のソース及び第1のドレイン領域を有し、前記第1のソース、及び第1のドレイン領域を隔てる領域に第1の絶縁膜を介して設けられた電位的にフローティング状態にある第1のフローティングゲート電極を有し、前記第1のフローティングゲート電極と第2の絶縁膜を介して容量結合する複数の第1の入力ゲート電極を有し、前記第1の入力ゲート電極の1つにソース電極が接続された第1のMOS型トランジスタを有し、前記第1のMOS型トランジスタには電位的にフローティング状態にある第2のフローティングゲート電極が設けられ、前記第2のフローティングゲート電極の延在部に接続部を介して電気的に接続された、第3のフローティングゲート電極を有し、前記第3のフローティングゲート電極部に電荷を出し入れするトンネル接合部を有するMOS型半導体装置において、少なくとも前記接続部の抵抗が前記トンネル接合の動作抵抗値よりも大きな値を有するよう構成されたことを特徴とする。

【0038】

【作用】本半導体装置は、少数の素子によってシナプス結合が構成でき、しかも電力消費が非常に少ないため、神経回路網の高集積化、低電力化が可能となる。さらに高精度のシナプス加重値の変更が可能となり、これによって初めて実用的なレベルのニューロンコンピュータチップを実現することができるのである。

【0039】

【実施例】（第1の実施例）本発明の第1の実施例を図1を用いて説明する。

【0040】図1（a）は、第1の実施例を示す回路図であり、図において101はニューロン回路であり、例

$$V_{TN} < V_{TP}$$

の条件に設定してある。

【0048】今、NMOS105、PMOS106に流

$$I_N = (1/2) \cdot \beta_N (V_{GS} - V_{TN})^2 \quad (2)$$

えば図11や図12に示した回路を用いればよい。より低消費電力化を求めるならば、図12の回路を用いた方が好ましい。102a～102dは、ニューロン回路の入力端子であり、例えば図12の回路であれば1206a～1206dに対応している。103はニューロンの出力信号を伝える配線であり、例えば図12に示したようなニューロン回路の出力端子1214に接続されており、そのニューロンが発火しているか否かに従って V_{DD} またはOVの電位を保持している。104は1つのニューロンの出力103と1つのニューロン101の入力102aを接続するシナプスの働きをする回路である。

【0041】次に104のシナプス回路について説明を行う。

【0042】105はNチャネルの ν MOS、106はPチャネルの ν MOSであり、それぞれのフローティングゲート107は電気的に接続されている。108は、 ν MOSのフローティングゲート107とコンデンサ C_a で容量結合している入力ゲートであり、前段のニューロンの出力線103に接続されている（その電位を V_a で表わす）。

【0043】109は、フローティングゲート112と、コンデンサ C_b によって容量結合する入力ゲートであり、シナプスの重み変更用の信号線110に接続されている。

【0044】一方、 ν MOSのフローティングゲート107は111の抵抗 R_c を介して、やはり電位的にフローティングのゲート112に接続されている。114は、例えば膜厚100Åの SiO_2 膜等の絶縁膜113を間に挟んで設けられた、電荷注入用の電極であり、その端子には必要に応じてバイアスが印加されるよう構成されている。

【0045】本回路の動作を説明するために、その主要部分の1つであるN- ν MOS105とP- ν MOS106を直列接続した回路の動作についてまず説明する。

【0046】図1（b）は、フローティングゲート107を入力端子 V_{in} として独立させて、説明用に描いた図面となっている。この回路の入出力特性、即ち V_{out} と V_{in} の関係について述べる。

【0047】一般に、MOS型トランジスタでは、キャリアの流れだす方の電極をソース、キャリアの流れ込む電極をドレインと呼んでいる。従ってNMOSでは電子の流れ出す低電位側がソースであり、高電位側がドレインと呼ばれる。また、PMOSでは、ホールが流れだす高電位側がソースであり、低電位側がドレインである。また、NMOS、PMOSの V_{in} からみたしきい値電圧はそれぞれ V_{TN} 、 V_{TP} であり、本実施例では、例えば、

$$(1)$$

れる電流をそれぞれ I_N 、 I_P とすると、いずれのトランジスタも飽和領域で動作しているので

$$I_P = (1/2) \cdot \beta_P (V_{GS} - V_{TP})^2 \quad (3)$$

となる。ここで、

$$\beta_N = (W/L)_N \mu_N C_{OX}$$

$$\beta_P = (W/L)_P \mu_P C_{OX}$$

$(W/L)_N$: NMOSのチャネル幅Wとチャネル長Lの比

$(W/L)_P$: PMOSのチャネル幅Wとチャネル長Lの比

μ_N : 電子のチャネル移動度

μ_P : ホールのチャネル移動度

図1(c)は、 I_N 、 I_P と V_{GS} の関係を示したものであり、 V_{in} に一定電圧が入力されているときは、 $I_N = I_P$ となる条件で回路が安定する。即ち、この時、

$$V_{GS} = V_{in} - V_{out} = V_T$$

となる。但し、 V_T は図より

【0049】

【数3】

$$V_T = \frac{\sqrt{\beta_R} V_{TN} + V_{TP}}{\sqrt{\beta_R} + 1} \quad \dots\dots (4)$$

で与えられる。

【0050】(1)式の条件 $V_{TN} < V_{TP}$ は、図1(c)で I_N と I_P の曲線が交点を持つための必要十分条件である。

【0051】従って、 V_{out} は

$$V_{out} = V_{in} - V_T \quad (5)$$

となる。

【0052】即ち、電圧ゲインが1の増幅器が実現できる。

【0053】今、 V_{in} が低い電位に変化したとする。即ち、

$$V_{out} > V_{in} - V_T$$

となったとすると、

$$V_{GS} = V_{in} - V_{out} < V_T$$

となる。この時、図1(c)より明らかなようにPMOSにはより多くの電流が流れ、NMOSに流れる電流は減少する。特に、 $V_{GS} < V_{TN}$ となるとNMOSはカットオフし、電流が全く流れなくなる。

【0054】このようにして、 C_{out} にたまった電荷は急速にPMOSにより放電し、 V_{out} は V_{in} の変化に追従し、再び $V_{out} = V_{in} - V_T$ となったところで落ちつくことになる。

【0055】逆に、 V_{in} が高電位側に変化し $V_{out} < V_{in} - V_T$ となったときには $V_{GS} = V_{in} - V_{out} > V_T$ となり、今度はPMOSがOFF状態に近づき、NMOS

には大きな電流がながれて C_{out} が急速に充電されるため、 V_{out} は上昇し再び $V_{out} = V_{in} - V_T$ となったところで落ちつく。

【0056】以上のように、 C_{out} の充・放電がそれぞれNMOSトランジスタとPMOSトランジスタのオンによって行われるため、高速の充放電が可能であり、高速で変化する入力信号に応答することができる。特に $\beta_N = \beta_P$ と設定してやるとNMOSとPMOSの電流駆動能力が等しくなり、充放電の速度が等しく、回路の高速化には特に有利である。

【0057】これは従来例にない大きな特徴である。例えば、図13の従来例では、 C_{out} の充放電時間はそれぞれ $R \cdot C_{out}$ 及び $R_x \cdot C_{out}$ で決まっており、Rや R_x を小さくしない限り充放電時間を小さくすることはできなかった。しかし、これらの抵抗値を小さくすると電圧ゲインが小さくなると共に消費電力が抵抗値に反比例して大きくなる等の問題があった。

【0058】本発明ではトランジスタの β_N 、 β_P を大きくすることにより幾らでも充放電の時間を短くすることが可能である。しかも、 β_N 、 β_P の値を大きくとっても電圧ゲインは常に1であり、小さくなることはない。しかも消費電力は図1(c)より定常的に流れる電流が、

【0059】

【数4】

$$I = (1/2) \cdot \beta_N \left(\frac{V_{TN} - V_{TP}}{\sqrt{\beta_R} + 1} \right)^2 \quad (6)$$

で与えられるので V_{TN} と V_{TP} の値をほぼ等しくすることで電流Iは幾らでも小さくすることができ、回路動作速度とは無関係に消費電力を一定の小さな値、あるいはほぼ0にすることが可能である。

【0060】以上述べたように、本発明の回路により、

従来のシナプス回路が持っていた消費電力が大きくなるという問題を解決することができた。

【0061】以上の説明では、図1(b)に示すNMOSの閾値 V_{TN} 、PMOSのしきい値 V_{TP} が、 $V_{TN} < V_{TP}$ となる場合について説明したが、これを

$$V_{TN} = V_{TP}$$

と設定してもよい。

【0062】こうすれば、(6)式より消費電力をゼロとすることができる。

【0063】あるいは、NMOSの閾値 V_{TN} 、PMOSのしきい値 V_{TP} は、

$$V_{TP} < V_{TN} \quad (7)$$

と設定してもよい。この場合は、図1(d)に示したように I_N 、 I_P の特性は交点を持たないため V_{GS} は

$$V_{TP} < V_{GS} < V_{TN}$$

の範囲で任意の値をとることになる。即ち、 V_{out} は

$$V_{in} - V_{TN} < V_{out} < V_{in} - V_{TP} \quad (8)$$

の間のいずれかの値に落ちつくことになるが一定値となる保障はない。つまり、上記の範囲で値の不確定性を持つことになる。回路動作上、この値の不確定性が許される範囲であれば(7)式の条件を採用してもよい。この場合、定常的に流れる電流は完全にゼロになるため、ほとんど電力消費のない回路が実現できる。

【0064】また V_{TP} と V_{TN} を $V_{TP} < V_{TN}$ の関係を満たしつつ、 $V_{TP} \doteq V_{TN}$ としてやれば、(8)式で表される V_{out} の値の不確定性は十分小さくすることができる。

【0065】また、 V_T の値は(4)式で与えられるが V_{TN} 、 V_{TP} 、 β_R の値を適宜設定して、 $V_T = 0$ とすることも可能である。例えば、 $\beta_R = 1$ 、 $V_{TN} = -1V$ 、 $V_{TP} = 1V$ とすれば、 $V_{out} = V_{in}$ となり、入力電圧と等しい値を出力側に取り出すことができる。また、この他

$V_T = 0$ とするために、 V_{TN} 、 V_{TP} 、 β_R に対しいかなる値を組み合わせてもよいことは言うまでもない。

【0066】このように設定すれば、常に $V_{in} = V_{out}$ となり、フローティングゲートの電位を直接 V_{out} に読み出すことができる。即ち、図1(a)のニューラルネットワークでは、フローティングゲート107の値がニューロン101の1つの入力端子102aへの入力信号となるのである。

【0067】次にフローティングゲート107の電位 ϕ_F の値を求める。図1(e)はその為に描いた回路図である。ここで $C_0 = C_{on} + C_{op}$ であり、 C_{on} 、 C_{op} はそれぞれフローティングゲート107とNMOS105、PMOS106の反転層の間の容量であり、それぞれのゲート酸化膜容量に等しい。

【0068】本回路の通常の動作状態においては、NMOS、PMOS共に反転層が形成されており、その電位はソースの電位、即ち V_{out} に等しい。特に(4)式の V_T を、 $V_T = 0$ とした場合には、定常状態では $V_{out} = \phi_F$ であり、 C_0 の両端にかかる電位差は0となる。つまり C_0 の効果は無視することができる。 C_E はフローティングゲート112と電荷注入ゲート114(その電位を V_E とする)との間の容量である。

【0069】フローティングゲートに蓄えられている電荷の量を Q_F とすると、 ϕ_F は、

【0070】

【数5】

$$\phi_F = \frac{C_a V_a + C_b V_b + C_E V_E + C_0 V_{out} + Q_F}{C_a + C_b + C_E + C_0} \quad \dots (9)$$

と表される。

【0071】通常のニュートラルネットワークの動作状態、即ち、学習によって重みを変える操作を行っていないときは、例えば $V_b = V_E = 0$ とする。そうすれば

(9)式より、

【0072】

【数6】

$$V_{out} = \frac{C_a V_a + Q_F}{C_a + C_b + C_E} \quad \dots (10)$$

となる。つまり、 C_0 の効果は無視できるのである。 V_a と V_{out} の関係を図1(f)に示す。 V_a は前段のニューロンの出力であり、ニューロンが発火しているか、していないかに応じて、それぞれ $V_a = V_{DD}$ 、もしくは $V_a = 0$ となる。即ち、前段のニューロンが発火し、 V

$$\phi_F = \frac{C_a V_a + C_0 V_{out} + Q_F}{C_a + C_b + C_E + C_0}$$

となり、 $\phi_F = V_{out}$ であることを考慮すると、

【0073】

【数7】

$V_a = V_{DD}$ となったときには、次段のニューロンへの入力電圧 V_{out} は、

【0074】

【数8】

$$V_{out} = \frac{C_a V_{DD} + Q_F}{C_{TOT}'} \quad \dots\dots (11)$$

$$= \frac{C_a}{C_{TOT}'} \left(1 + \frac{Q_F}{C_a V_{DD}} \right) \cdot V_{DD} \quad \dots\dots (11)$$

$$C_{TOT}' = C_a + C_b + C_E \quad \dots\dots (12)$$

となり、104のシナプス回路で発生する重みwは、

$$w = \frac{C_a}{C_{TOT}'} \left(1 + \frac{Q_F}{C_a V_{DD}} \right) \quad \dots\dots (13)$$

となる。

【0076】一例として、 $C_a : C_b : C_E = 8 : 1 : 1$ と設計すると、 $C_a / C_{TOT}' = 0.8$ となり、 $Q_F = 0$ で、 $w = 0.8$ となる。wの値を変更するには、 Q_F の値を変化させればよく、これには電荷注入用の電極114より、絶縁膜を介してフローティングゲート112に電子をFowler-Nordheim Tunnelingにより注入することにより、あるいは、フローティングゲートから電極114に電子を放出することにより行う。

【0077】例えば電子注入を行って、 $Q_F < 0$ となれば、図1(f)に示した様に $V_a = V_{DD}$ のときの出力レベルが下がり(13)式で表されるシナプスの加重値が減少する。逆に、電子を放出して $Q_F > 0$ とするとwは大きくなる。特に、 $Q_F = (C_b + C_E) V_{DD}$ とすれば $w = 1$ となり、最大の加重値を得ることができる。

【0078】このときは、 $V_a = 0$ 、即ち前段のニューロンが発火していないときにも、シナプス104は、

【0079】

【数10】

$$\frac{C_b + C_E}{C_a + C_b + C_E} V_{DD}$$

(今の例では、 $0.2 V_{DD}$ に等しい)の出力を持つことになる。つまり、このシナプスを介して結合している相

$$\phi_F' = V_{PP} \frac{C_b}{C_E + C_b} = \frac{1}{2} V_{PP} \quad \dots\dots (14)$$

となる。

【0084】100ÅのSiO₂薄膜にパルス印加とともに約10Vの電圧がかかり、トンネリングによって電子が電極114よりフローティングゲート112に注入される。注入による電子数の時間変化の様子を模式的に示したのが図1(h)であるが、パルス印加の初期に大きく変化するが、その後は、ほとんど変化していないこ

$C_a V_{DD}$ のニューロンに常に正のバイアス进行ける結果となりそのニューロンを発火しやすくする効果がある。しかし、従って、もしこのような動作が許されない回路であれば Q_F は常に負として用いるべきであり、 $w \leq 0.8$ の範囲の値とすべきことになる。

【0080】 $Q_F > (C_b + C_E) V_{DD}$ とした場合には、加重値は $w = 1$ のままであり、この $V_a = 0$ での出力値(オフセット)だけが電荷量に比例して大きくなって行く。wの最大値として1をとり、しかも $V_a = 0$ でのオフセットをなくすシナプス回路の構成法は、本発明の第4、第5の実施例に示す。

【0081】次に、電子を注入する方法について説明する。

【0082】まず、第1の方法は、 $V_a = V_E = 0$ とし、 V_b 端子109に図1(g)に示した様なパルスを印加する。112と114の間に、例えば100Åの厚さの熱酸化膜(SiO₂)が設けられているとすると、 V_{PP} は約20V程度とすればよい。パルス巾 τ を $R_C C_a$ より十分小となるように接続部111の抵抗値 R_C を設定しておく、フローティングゲート107の電位 ϕ_F は V_b に印加されるパルスによって変化せず、フローティングゲート112の電位(ϕ_F' と表す)のみが変化し、

【0083】

【数11】

$$\frac{1}{2} V_{PP} \quad \dots\dots (14)$$

とが分る。これは、電子の注入によりフローティングゲートの電位 ϕ_F' が下がり、SiO₂膜にかかっていた電圧が減少するためである。

【0085】Fowler-Nordheim Tunnelingによる電流Iは、酸化膜両端にかかる電圧Vに対し、
 $\propto V^2 \exp(-b/V)$
 の依存性をもち、Vの減少とともに指数関数的に電流が

減少するからである。これは従来例に関し、図15

(b)で説明したのと同じ理由である。従って、パルス巾 τ を Δn の立上りの時間より十分大きくとっておけば、一回の注入操作による電荷注入量をほぼ一定の値(Δn_0)とすることができる。

【0086】一回の注入後、 $t = RC_a$ 程度の時間を経

$$\Delta w = - \frac{1}{C_{TOT}} \frac{q \Delta n_0}{V_{DD}} \dots\dots (15)$$

となり重みが減少する。逆に重みを増加させる。すなわち、電子を放出させるときは $V_b = V_a = 0$ とし、 V_E に図1(h)に示したパルス印加してやればよい。こ

過すると112に注入された電荷は107の領域に流れ込み、十分時間が経過した後は $\phi_F' = \phi_F$ となる。

このとき w の変化は(13)式より、

【0087】

【数12】

のとき w の変化は1回のパルスに対し

【0088】

【数13】

$$\Delta w = \frac{1}{C_{TOT}} \frac{q \Delta n_0}{V_{DD}} \dots\dots (16)$$

となり(但し $C_E = C_b$ の場合)、重みが増強される。 $\Delta n_0'$ は放出された電子の数であり、一般に Δn_0 と $\Delta n_0'$ は等しくない。しかし例えばフローティングゲート112として N^+ ポリシリコン、 V_E 端子(114)としてシリコンの N^+ 拡散層を用いればほぼ等しくなる。あるいは、必要に応じて注入時・放出時のパルスの高さをかえることにより、 $\Delta n = \Delta n_0$ としてもよい。

【0089】第1回目のパルス印加後、 RC_a 程度以上の時間の経過後、再び同様のパルス印加するとやはり(15)(16)式に従って w を連続的に変化させることができる。パルスの数を N_P 、書き換え前のフローティングゲート電荷の量を Q_{F0} とすると、

【0090】

【数14】

$$w = \frac{C_a}{C_{TOT}} \left(1 + \frac{Q_{F0} \mp N_P \Delta n_0}{C_{TOT} V_{DD}} \right) \dots\dots (17)$$

ここで \pm はそれぞれ電子の注入、放出に対応し、 $\Delta n_0 = \Delta n_0'$ とした。

【0091】図1(i)は上で述べた重み変化を実験的に確かめるために試作したテストデバイスの構造を模式的に示したものであり、単体のトランジスタ115のフローティングゲート107に、図1(a)のシナプス回路と同様の原理で電荷注入を行えるようにしたものである。

【0092】実験結果を図1(j)に示す。パルスの数とともにほぼ直線的に V_{TH} が変化していることが分かる。即ち、1パルス毎にほぼ一定量の電荷の注入もしくは放出が制御よく行われているのである。ここでパルス電圧の大きさは19.5V、パルス巾は10msec、各パルスと各パルスの間隔は100msecである。

【0093】このように、制御性よく電荷注入を行えるのは、フローティングゲート107とフローティングゲート112の間に高抵抗の接続部111を設けた結果であり、これが本発明の大きな特徴である。即ち、各パルス毎に一定量の電荷をフローティングゲート112に先ず注入する。次に、この電荷を高抵抗接続部111を介してシナプス回路104の主要ゲート電極であるフローティングゲート107にゆっくりと流し込んでやるので

【0093】一回のパルスで飽和するのは、112の部分のみであり、全体(112と107)が同時に飽和することは決してないのである。

【0094】その結果、各パルス毎にほぼ一定量の電荷を注入することが可能となった。つまり本発明によりはじめて、フローティングゲートへの電荷の注入あるいはフローティングゲート103の電荷の放出をいづれもパルスの個数によって正確に制御できるようになり、このことによりシナプスの加重値を精密に制御することが可能となった。

【0095】また、104の2つの ν MOSを組合せた構造により、高速に応答し、且つ、パワー消費の極めて少ないシナプス結合が実現できたのである。しかも全体として、たった2つのMOS型トランジスタにより一個のシナプスが形成でき、従来の方法にくらべ大幅な回路の簡略化が達成できたのである。これにより大規模なニューラルネットワークを容易に実現できるようになったのである。

【0096】次に、104のシナプス回路のフローティングゲートに電子を注入もしくは、放出する第2の方法について説明する。

【0097】これは、 V_a 、 V_b に独立の電圧を加え、

特定のシナプスのみ選択的に加重値の変更を行う方法であり、ニューラルネットワークに学習を行わせる際に非常に重要な方式である。

【0098】一例として、 $C_a : C_b : C_E = 8 : 1 : 1$ の場合を説明する。まず加重値の変更を行うシナプスでは $V_a = 6.25 \text{ V}$ とし、それ以外のシナプスでは $V_a = 0$ とする。これには信号線103の所定のもののみ 6.25 V の電位を与えればよい。 $V_b = V_E = 0$ とするとフローティングゲートの電位は、(10)式より $Q_F = 0$ として

$$\phi_F = (8V_a / 10)$$

$$\Delta \phi_F' = V_{PP} \frac{C_b}{C_E + C_b} = \frac{1}{2} V_{PP}$$

となる。

【0101】ここでは $\Delta \phi_F' = 5 \text{ V}$ である。従って $V_a = 6.25 \text{ V}$ の加えられているシナプスでは $\Delta \phi_F' = 10 \text{ V}$ となり、 $V_a = 0 \text{ V}$ のシナプスでは $\Delta \phi_F' = 5 \text{ V}$ となる。つまり $V_a = 6.25 \text{ V}$ 、 $V_b = 10 \text{ V}$ (パルス) のシナプスでのみ加重の変更が行えるのである。

【0102】以上は電子注入により加重を減少させる場合について述べたが、電子放出により加重を増加させる場合も同様に行える。まず加重値を増加させるシナプスでは、 $V_a = 3.75 \text{ V}$ とし、それ以外のシナプスでは $V_a = 10 \text{ V}$ とする。これには、信号線103の所定のもののみ 3.75 V の電位を与え、それ以外には 10 V の電位を与えればよい。 $V_b = 10 \text{ V}$ 、 $V_E = 10 \text{ V}$ とするとフローティングゲートの電位は

$$\phi_F = (8V_a + 20) / 10$$

となるので、 $V_a = 3.75 \text{ V}$ では、 $\phi_F = 5 \text{ V}$ 、 $V_a = 10 \text{ V}$ では $\phi_F = 10 \text{ V}$ であり、トンネル酸化膜にかかる電圧 V は、 $V_a = 3.75 \text{ V}$ 、 $V_a = 10 \text{ V}$ のそれぞれの場合に対し $V = 5 \text{ V}$ 、 $V = 0 \text{ V}$ となるため、 100 \AA のトンネル酸化膜ではトンネル電流は流れない。

【0103】次に、加重増加を行うシナプスにつながっている信号線110についてのみ 10 V のバイアス値をパルスのに 0 V に落とす。この負のパルスによるフローティングゲート112の電位 ϕ_F' の変化分 $\Delta \phi_F'$ は、(14)式より $\Delta \phi_F' = -5 \text{ V}$ となる。

【0104】従って、 $V_a = 3.75 \text{ V}$ のシナプスでは $\phi_F' = 0$ となり、トンネル酸化膜にかかる電圧は 10 V となり、トンネリングによって電子がフローティングゲートから114の電極にぬける。従って、 Q_F は正の値で増加し、シナプスの加重値が増加する。一方、 $V_a = 10 \text{ V}$ のシナプスでは $\phi_F' = 5 \text{ V}$ となるためトンネリングは生じない。

【0105】即ち、 $V_a = 3.75 \text{ V}$ 、 $V_b = 0 \text{ V}$ としたシナプスにおいてのみ選択的に加重値の変更が行われ

となるので、 $V_a = 6.25 \text{ V}$ では、 $\phi_F = 5 \text{ V}$ となる。酸化膜両端の電圧が 5 V の条件では、 100 \AA のトンネル酸化膜で、トンネリングは生じず、従ってキャリアの注入による重みの変化は生じない。

【0099】次に電子注入を行うシナプスにつながっている信号線110にのみ、 10 V のパルス電圧を印加する。即ち、 V_b に 10 V のパルスを加えるのである。このとき、トンネル注入部のフローティングゲート112の電位 ϕ_F' の変化分 $\Delta \phi_F'$ は、(14)式より

$$\text{【0100】}$$

$$\text{【数15】}$$

るのである。

【0106】以上の説明で用いた電圧の設定値や組合せはあくまで一例であり、方式の原理を説明するためのものである。従っていかなるバイアスのかけ方を用いてもよいことは言うまでもない。もちろん、 V_E 端子114にパルスを印加する方式をとってもよいことは言うまでもない。

【0107】本発明のシナプスの回路を用いることにより、シナプスの加重変更に必要なパルス電圧の低減も実現できた。

【0108】従来は選択的変更を行うためには、 20 V という高電圧のパルスが必要であったが、電荷注入部のフローティングゲートを接続の抵抗 R_C によって切りはなしたため、制御ゲート(109)の制御性がよくなり、最大 10 V で選択的変更が可能となったのである。大規模なニューラルネットワーク構成に極めて有利な特徴である。

【0109】接続部 R_C の抵抗値に関しては次のように決めればよい。トンネリングによって、電子がフローティングゲート112に注入される時間を τ_p と表す。 τ_p としては、例えば、図1(h)で $\Delta n(t) = 0.9 \Delta n_0$ となる時間を目安としてもよい。あるいは、トンネル電流 $i(t)$ がパルス印加直後の最大値に対し例えば $1/10$ となる時間と定義してもよい。

【0110】また一方、注入部のフローティングゲート112に注入された電荷が、トランジスタ部のフローティングゲート107に流れ込むのに要する時間は、時定数 $R_C C_a$ 程度であるから、 $R_C C \gg \tau_p$ となるように R_C を決めてやればよい。トンネル接合部の動作抵抗 r_t は、

$$r_t \equiv (di/dV)^{-1} \quad (18)$$

で定義される。

【0111】 i と V は、Fowler Nordheimの式

$$i = AV^2 \exp(-b/V) \quad (19)$$

と表されるから、

$$(\gamma_T)^{-1} = (2AV + Ab) \exp(-b/V) \quad (20)$$

となる。ここでVはトンネル部の絶縁膜両端にかかる電圧であり、例えば $V_E = 0$, $V_b = V_{PP}$ として電子注入

を行う際には、 $V = \phi_F$ となり、

$$\phi_F(t) = \frac{b}{\ln[\exp\{b/\phi_F(0)\} + Abt/C_{TOT}]} \quad (21)$$

と表される。ここで、 $C_{TOT} = C_E + C_b$, A, bは(20)式中の定数で、絶縁膜材料や注入電極材料等で決まる定数である。また $\phi_F(0)$ は、 V_{PP} の大きさのプログラム電圧をステップ関数状に V_P 端子に印加した

直後のフローティングゲートの電位であり

【0112】

【数16】

$$\phi_F(0) = \frac{C_b \cdot V_{PP}}{C_E + C_b} \quad (22)$$

で与えられる。tは、プログラム電圧印加後の時間である。

時間とともに減少するが、これとともに r_T は(20)式より極端に大きくなることが分かる。 R_C を決める条件は、 $0 < t < \tau_P$ の時間において

$$\gamma_T(C_E + C_b) \ll R_C C_a \quad (23)$$

が成立つようにとればよい。

【0114】即ち一般的に言って

(24)

$$r_T < R_C$$

であることが必要条件であり、望ましくは $\gamma_T \leq R_C$ であるが、必要に応じていかなる値をとってもよいことは明らかである。

リコンを用いてもよいことはいうまでもない。

【0115】通常 $C_a > C_E + C_b$ と設定することが多いので、(24)式を満たしておれば、ほぼ(23)式の関係を満たすことになる。 R_C を実現する方法としては、例えば112, 107として、 N^+ ポリシリコンを用い、接続部111には、高抵抗ポリシリコンを用いてもよい。あるいは112, 107をすべて高抵抗ポリシリコンを用いてもよいことは言うまでもない。

【0120】また V_a に前段のニューロンの出力が入力され、シナプス回路として動作しているときも $V_c = 0$ として203のスイッチをOFFしてやればRC時定数で電荷が変化し、動作中に重みが増加することもなくなり動作精度を上げることができる。

【0116】(第2の実施例)図2は、本発明の第2の実施例を示す図面であり、フローティングゲート部の構造のみを模式的に示してある。

【0121】(第3の実施例、第4の実施例)次に本発明の第3、第4の実施例をそれぞれ図3(a)、図3(b)に示す。

【0117】201, 202は、例えば N^+ ポリシリコンで形成したフローティングゲートであり、図1(e)の112, 107にそれぞれ対応している。203はP型のポリシリコンであり、その上部にはゲート絶縁膜204を介して電極205が設けられている。

【0122】第1の実施例(図1(a))ではNチャンネル型とPチャンネル型の2つの ν MOS105, 106が互いにフローティングゲートを共用していたが、これは片方のみを ν MOSとし、他方を通常のMOSトランジスタとしてもよい。

【0118】即ち、接続部(図1(e)の111に相当)がポリシリコンの薄膜トランジスタ(TFT)で形成されているのである。トンネル注入時には $V_c = 0$ としてこの接合部のTFTをOFFとするが、注入終了後には $V_c = V_{DD}$ としてTFTをONし、すみやかに電荷を201より202へ移動させるのである。1回のパルス印加後、次のパルス印加までの間に十分長い時間をとる必要がないため、シナプスの加重変更が迅速に行えるという特徴がある。

【0123】図3(a)の第3の実施例は、NMOS301のみを ν MOSとした場合であり、PMOS302は通常のMOSトランジスタとなっている。そのゲート電極303はたとえば制御信号 ϕ_c でコントロールする。 $\phi_c = 0$ のときはPMOS302がONし、出力端子304を0Vにリセットする。シナプス回路として働かせるときは $\phi_c = V_{DD}$ とする。従って ϕ_c は V_a と同じ信号を用いてもよい。即ち、303は103に接続してもよい。

【0119】本実施例は N^+ ポリシリコンとP型ポリシリコンの組合せで説明したが、これは例えば N^+ ポリシリコンとN型ポリシリコン、あるいは201, 202に P^+ ポリシリコン、203にN型あるいはP型のポリシ

【0124】図3(b)の第4の実施例は、PMOS305のみを ν MOSとした場合を示している。これらの方式は、P-wellあるいはN-wellのバイアス電圧の加え方が簡単になるというメリットをもっている。P-well方式では図3(a)の方式が、N-well方式では図3(b)の方式が適している。つまりトランジスタ301, 305等の形成されているウェルの電位をそれぞれ $V_{SS}(0V)$, V_{DD} に固定するのでは

なく、 V_{out} 端子と接続することにより、トランジスタのバックバイアス効果による閾電圧の増加を防ぐことができるのである。

【0125】（第5の実施例、第6の実施例）次に本発明の第5、第6の実施例を図4（a）、図4（b）にそれぞれ示す。図4（a）の第5の実施例では、加重値のデータをたくわえているNチャネル型の ν MOS 401がNMOS 402とPMOS 403にはさまれた形の回路となっている。これらの2つの通常のMOSトランジスタのゲートは同一の制御信号 ϕ_c でコントロールされている。 $\phi_c = 0$ はリセットモードであり出力端子404は通にNMOS 402で接地され、 $V_{out} = 0$ となっている。

【0126】ここで本発明の実施例で重要なことは、PMOS 403がOFFしていることであり、リセットモードで直流電流が流れることがなく消費電力低減に非常に有利となっている。

【0127】図3（a）、（b）の回路では、 ν MOS 301あるいは305がフローティングゲートにそれぞれ正もしくは負の電荷が注入され、デプレションモードとなっている場合にはリセットモードで貫通電流が流れ消費電力が増大してしまう。従って、第3、第4の実施例では、それぞれ正もしくは負の電荷を保持することは好ましくなかった。さらに図1（a）の回路では、 $Q_F > 0$ のとき、図1（f）で述べた $V_a = 0$ でのオフセットが生じる問題があった。これらすべてがこの第5の実施例では解決されているのである。

【0128】 Q_F を正で大きな値とし、 $w = 1$ としてもオフセットが生じないし、またリセットモードで貫通電流が流れることは、いかなる場合にも発生しないという大きな特徴をもっている。

【0129】 ϕ_c はリセットモードでは V_{DD} 、シナプスとして働かせるときは0Vとなる信号を用いればよい。また ϕ_c は、独立の信号ではなく、前段のニューロンの出力を反転させた信号線につないでおいてもよい。ニューロンがクロックと同期してクロックがHIGHのときのみ0又は1の出力を出し、それ以外のときには0を出力しているのであればこの方式が簡単である。こうした場合、 V_a は必ずしもニューロンの出力線103に接続する必要はなく、一定の電位を与えておいてもよい。たとえば常に V_{DD} に固定しておいてもよい。

【0130】図4（b）に示した第6の実施例は、第5の実施例のNチャネル ν MOS 401の代わりにPチャネル ν MOS 405を用いたものである。いずれの構成を用いてもよいことは言うまでもない。

【0131】上記第5、第6の実施例はそれぞれP-well、N-wellに用いた場合には、 ν MOS 401、405の形成されているウェルの電位を V_{out} 端子404、406と接続することができ、これらのトランジスタのバックバイアス効果による閾値の変動をなくす

ことができるため、シナプスの重みをさらに正確に決定できる。

【0132】（第7の実施例）図5は本発明の第7の実施例を示す半導体装置の断面を表す図面であり、例えば図4（a）の回路をN-wellプロセスで実現させたときに、Nチャネル ν MOS 401のバックバイアス効果を有効に防止する方法を示している。

【0133】P型のシリコン基板501は通常0VにアースされているためNチャネル ν MOS 502（図4（a）の401）の基板503は、 V_{out} 端子504と接続できないが、本実施例では、503と501の間にN型のバッファ層505をはさむことにより解決している。

【0134】505は503と同変位に接続されている為、両者の間に電流は流れない。また505と501は常に逆バイアス状態となっているため電流は流れないのである。

【0135】本実施例は、Nチャネル ν MOSの場合について述べたが、例えば図4（b）のPチャネル ν MOS 405にも同様の構成が用いられることは言うまでもない。この場合は、 V_{DD} にバイアスされたN型基板中のN型のウェルをP型のバッファ層が囲む形となる。

【0136】本発明の第1の実施例、図1（a）の回路では、例えばPウェル方式で実現するには、Nチャネル ν MOS 105にこのような二重ウェル方式を採用し、Nチャネル ν MOS 105、Pチャネル ν MOS 106共にそのウェルの電位を V_{out} に接続することができる。

【0137】Nウェル方式の場合はPMOSに二重ウェルを採用することになる。NMOS、PMOSがそれぞれ互いに分離された絶縁基板上的のシリコンアイランド内に形成されるSOIデバイスでは、このようなウェル構造が不要であることは言うまでもない。

【0138】さらに基板を必ずしも V_{out} と接続する必要はない。基板バイアス効果が入った状態でも正確に動作させることは可能であり、何ら問題の生じることはない。

【0139】（第8の実施例）図6（a）は本発明の第8の実施例を示す回路の構成図面である。

【0140】 ν_I^1 、 ν_I^2 、 \dots 、 ν_I^n は第1層のn個のニューロン群を表しており各々は図1（a）の101と同様の回路を有している。 ν_{II}^1 、 ν_{II}^2 、 \dots 、 ν_{II}^m 、及び ν_{III}^1 、 ν_{III}^2 、 \dots 、 ν_{III}^m はそれぞれ第2層のn個のニューロン群、第3層のm個のニューロン群をあらわしている。各ニューロンはすべて101と同様の構造を有している。 O_I^1 、 O_I^2 、 \dots 、 O_I^n は第1層のニューロンからの出力線であり、図1（a）の103の信号線に対応している。

【0141】また I_{II}^1 、 I_{II}^2 、 \dots 、 I_{II}^m は第2層のニューロン群への入力線であり、図1（a）の102

a ~ 102d のフローティングゲートに相当している。 O_{II}^1 、 O_{II}^2 、 O_{II}^3 、…、 O_{II}^n は第2層のニューロン群の出力線であり、601a、601b等の配線を介して第3層のニューロン群の入力線と結合している。また602a、602b、等の配線により、第1層ニューロン群の出力線と平行の配線603a、603b等に接続されており、フィードバックループを形成している。

【0142】604a、604b等はシナプス結合回路であり、例えば図1(a)の104に対応している。

【0143】本実施例では、3層のニューロン群からなる神経回路網のレイアウトを示したものであるが、本発明によると、2次元的に規則正しく配列することにより神経回路を構成することができ、LSI化にとって極めて有利であることが分る。

【0144】図6(b)は、同図(a)を簡略化して示したものである。605の $v_I(n)$ は第1層のニューロン群を表しておりnはn個のニューロンからなることを意味している。606の $S(2n, n)$ は2n個の出力(第1層のn個の出力と第2層のn個の出力がフィードバックされたもの)とn個の第2層への入力とを結びつけるマトリクス状に配置されたシナプス群を表しており、シナプスマトリクスと呼ぶ。607の矢印は信号の流れを表している。608の矢印はn個の出力がフィードバックされていることを表している。このような表記法を用いれば、もっと複雑な神経回路網をも簡単に表現することができる。

【0145】(第9の実施例)図7(a)は、図6(b)の表記法を用いて表現した4層の神経回路網であり、本発明の第9の実施例である。ここで $I(n)$ は入力バッファ層であり、同図(b)に示したようにn個の増幅器からなっており、「0」または「1」の2進信号の入力 I_1 、 I_2 、…、 I_n をそれぞれ0V及び V_{DD} のレベルとして出力する回路である。図7(a)より明らかなように、フィードバックを有する4層の神経回路網が規則正しい2次元的な配列によって実現されている。人間の脳は6層のニューロン群から構成されているといわれているが、同様の配列を平面的にくり返すことにより、何層のニューロンを含む神経回路網でも容易に構成できる。すなわち本発明の半導体装置は、ニューロンコンピュータ実現に非常に有利な特徴を有している。

【0146】(第10の実施例)次に本発明の第10及び第11の実施例をそれぞれ図8(a)、(b)に示す。

【0147】図8(a)において801は図1(a)101で表されるニューロン回路である。802はそのフローティングゲートであり、803a、803b、803cは、例えば図1(a)の104のようなシナプス回路の出力線である。

【0148】本実施例では、フローティングゲート80

2がスイッチングトランジスタ803を介して接地されている。これは安定したニューロン動作を実現するためであり、例えばホットエレクトロン注入等によりフローティングゲート内に注入された電荷を、 ϕ_m の信号でトランジスタ803を導通させ、イニシャライズする機能を装備したものである。回路の信頼性を向上させる上で有効な方法である。

【0149】(第11の実施例)さらに図8(b)の第11の実施例は、トランジスタ803を介して信号電圧 V_m に接続されている。 V_m の値を適宜設定することにより、所定の電荷 Q_F をフローティングゲート802に与え、ニューロンの発火する閾値を任意の値に設定することができる。ニューロンの機能変更を自在に行うことができ、さらに高度なニューラルネットワークの構成が可能である。

【0150】シナプス結合の大きさを決める要因として、例えば図1(a)のコンデンサ C_a の大きさも重要である。 C_a は大きい程ノイズマージンが大きくなるが、その為には電極108の面積を大きくしなければならない。面積を増やすことなく大きな容量を実現するためには、 Ta_2O_5 、 Al_2O_3 等の高誘電体材料を電極108とフローティングゲート107の間の絶縁膜として用いればよい。

【0151】また、ニューロン回路101においても入力ゲートとフローティングゲート間の絶縁膜はやはり高誘電体材料を用いるのがよい。これによりシナプス出力とニューロン入力との結合部の面積を小さくすることができる。フローティングゲートの電位変化は通常入力ゲートの容量 C_I とフローティングゲートと基板間の結合容量 C_0 の比 C_I/C_0 で決まる。従って C_I は大きく、 C_0 は小さくすることが望ましい。

【0152】即ちコントロールゲートとフローティングゲート間には高誘電体絶縁膜を用い、フローティングゲートと基板間には例えば従来の SiO_2 のように比較的誘電率の小さな材料を用いるのがよい。

【0153】

【発明の効果】本発明によれば、少数の素子によってシナプス結合が構成でき、しかも電力消費が非常に少ないため、神経回路網の高集積化、低電力化が可能となる。さらに高精度のシナプス加重値の変更が可能となり、これによって初めて実用的なレベルのニューロンコンピュータチップを実現することができるのである。

【図面の簡単な説明】

【図1(a)】第1の実施例を示す回路図。

【図1(b)】フローティングゲート107を入力端子 V_{in} として独立させた回路図。

【図1(c)】 I_N 、 I_P と V_{GS} の関係を示すグラフ。

【図1(d)】 I_N 、 I_P の特性を示すグラフ。

【図1(e)】フローティングゲート107の電位 ϕ_F の値を求めるための回路図。

【図1 (f)】 V_a と V_{out} の関係を示すグラフ。
 【図1 (g)】 V_b 端子に印加するパルス図。
 【図1 (h)】 フローティングゲート112に注入される電子数の時間変化の様子を模式的に示したグラフ。
 【図1 (i)】 重み変化を実験的に確かめるために試作したテストデバイスの構造を模式的に示した図。
 【図1 (j)】 図1 (i) のデバイスを用いた実験の実験結果を示すグラフ。
 【図2】 第2の実施例を示し、フローティングゲート部の構造を模式的に示した図。
 【図3】 図3 (a) は第3の実施例を示す回路図、図3 (b) は、第4の実施例を示す回路図。
 【図4】 図4 (a) は第5の実施例を示す回路図、図4 (b) は第6の実施例を示す回路図。
 【図5】 第7の実施例を示す半導体装置の断面概念図。
 【図6 (a)】 第8の実施例を示す回路の構成図。
 【図6 (b)】 図6 (a) を簡略化して示した図。
 【図7】 第9の実施例に係り、図7 (a) は図6 (b) の表記法を用いて表現した4層の神経回路網を示し、図7 (b) は n 個の増幅器からなる入力バッファ層1 (n) を示す。
 【図8】 図8 (a) は第10の実施例を示し、図8 (b) は第11の実施例を示す。
 【図9】 脳の基本単位の構成をモデル化した図。
 【図10】 図10 (a) は1つの神経細胞、すなわち1個のニューロンの機能を説明する概念図、図10 (b) は Z と V_{out} の関係を表したグラフ。
 【図11 (a)】 ν MOS構造の一例を示す簡略化した概念図。
 【図11 (b)】 図11 (a) の構造をさらに簡略化した図。
 【図11 (c)】 図11 (a) のニューロン素子を用いたインバータ回路図。
 【図11 (d)】 図11 (c) の回路における V_{out} 、 V_{in} を Z の関数として示したグラフ。
 【図12 (a)】 CMOSニューロンゲートの断面構造を模式的に表した図。
 【図12 (b)】 1個のニューロン回路を構成示す回路図。
 【図13】 従来技術による ν MOSトランジスタを用いたシナプス結合も含むニューロン回路の基本構成の一例を示す回路図。
 【図14】 図14 (a) は可変抵抗の実現方法の一例を示し回路図、図14 (b) は V_{GG} の値を制御する一例を示す回路図。
 【図15】 図15 (a) はトンネル接合を有する E^2P ROMセルの閾電圧 (V_{TH}) を、データ書き込み用のパルスの数の関数として示したグラフであり、図15 (b) は正のプログラム電圧をステップ関数的に印加したときの、フローティングゲートに注入される電子の数

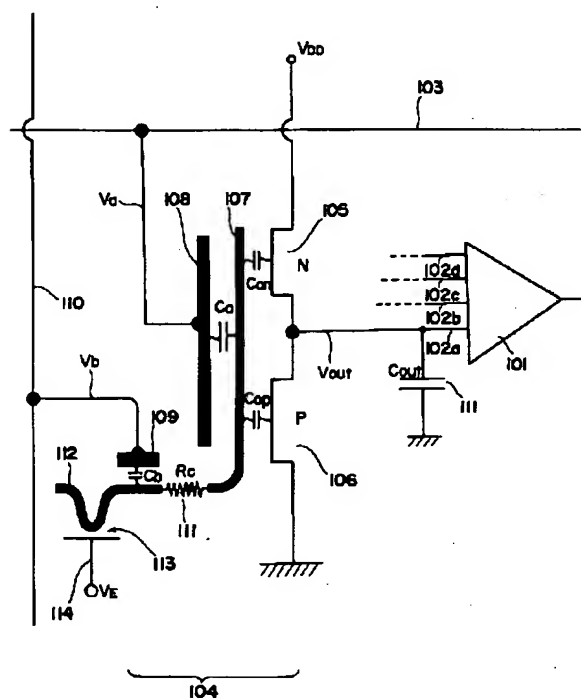
(n) の時間変化の様子を示したグラフ。

【符号の説明】

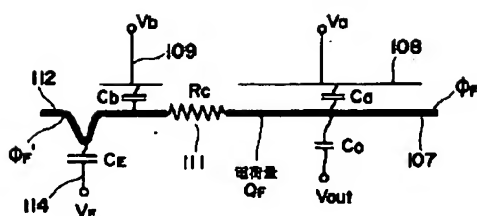
101 ニューロン回路、
 102 a ~ 102 d ニューロン回路の入力端子、
 103 出力線 (ニューロンの出力信号を伝える配線)、
 104 シナプスの働きをする回路 (シナプス回路)、
 105 Nチャネルの ν MOS、
 106 Pチャネルの ν MOS、
 107 フローティングゲート、
 108 入力ゲート、
 109 入力ゲート (制御ゲート、端子)、
 110 シナプスの重み変更用の信号線、
 111 抵抗 (接続部)、
 112 フローティングゲート、
 113 絶縁膜、
 114 荷注入用の電極、
 201, 202 フローティングゲート、
 203 P型のポリシリコン、
 204 ゲート絶縁膜、
 205 電極、
 301 トランジスタ、
 302 PMOS、
 304 出力端子、
 305 PMOS、
 401 チャネル型の ν MOS、
 402 NMOS、
 403 PMOS、
 404 出力端子、
 405 ν MOS、
 406 V_{out} 端子、
 501 P型のシリコン基板、
 502 Nチャネル ν MOS、
 503 基板、
 504 V_{out} 端子、
 505 バッファ層、
 601 a, 601 b 配線、
 602 a, 602 b 配線、
 603 a, 603 b 配線、
 604 a, 604 b シナプス結合回路、
 605 第1層のニューロン群、
 801 ニューロン回路、
 802 フローティングゲート、
 803 a, 803 b, 803 c シナプス回路の出力線、
 802 フローティングゲート、
 803 トランジスタ、
 901 a, 901 b, 901 c ニューロン、
 902 a, 902 b, 902 c 神経繊維、
 903 a, 903 b, 903 c シナプス結合、

1101 シリコン基板、
 1102, 1103 ソース及びドレイン、
 1104 ゲート絶縁膜、
 1106 フローティングゲート、
 1107 絶縁膜、
 1108 入力ゲート、
 1110, 1111 インバータを構成するための抵抗、
 1112 NMOSトランジスタ、
 1201 シリコン基板、
 1202 ウェル、
 1203a ソース、
 1203b ドレイン、
 1204a ソース、
 1204b ドレイン、
 1205 フローティングゲート、
 1206a～d 入力ゲートの電極、

【図1a】

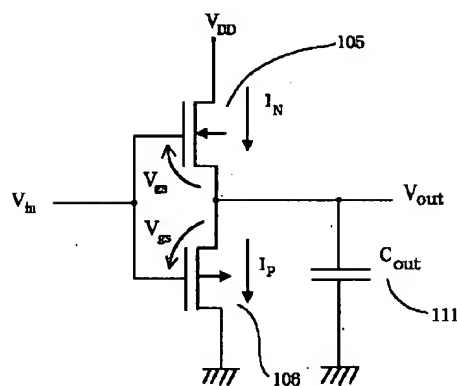


【図1e】

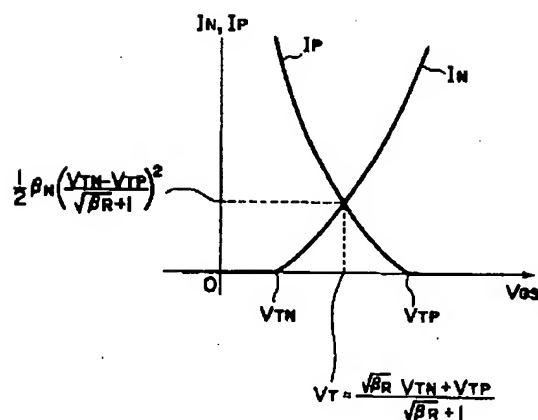


1207, 1208 絶縁膜、
 1209 フィールド酸化膜、
 1210 CMOSニューロングート、
 1211 CMOSのインバータ、
 1212 NMOSトランジスタ
 1213 PMOSのトランジスタ、
 1214 ニューロン回路の出力端子、
 1301 ニューロン回路、
 1302 他のニューロンの出力信号を伝える配線、
 1303 シナプス結合回路、
 1304 NMOSトランジスタ、
 1305 ゲート電極、
 1306 ソース、
 1307 結線、
 1401 MOSトランジスタ、
 1402 バイナリーカウンタ、
 1403 D/Aコンバータ。

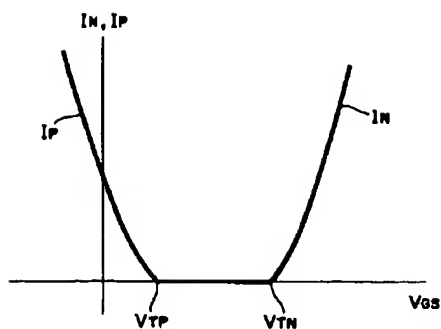
【図1b】



【図1c】

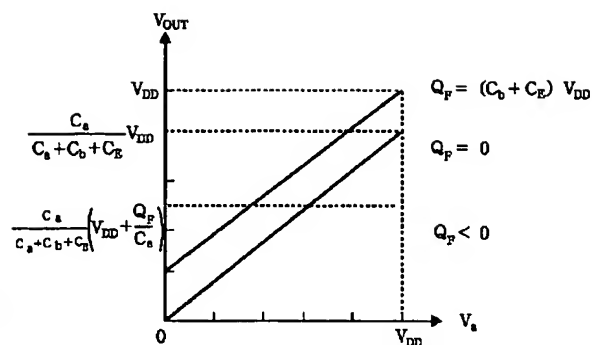


【図 1 d】



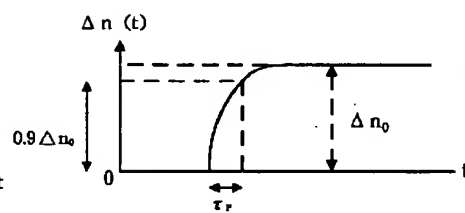
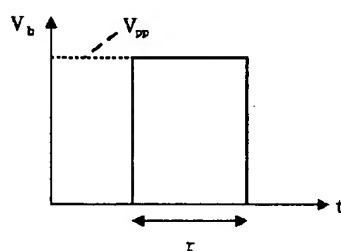
【図 1 g】

【図 1 f】

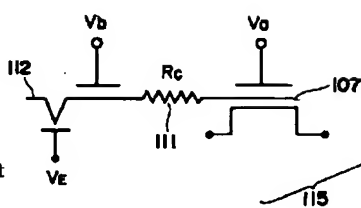


【図 1 h】

【図 1 i】

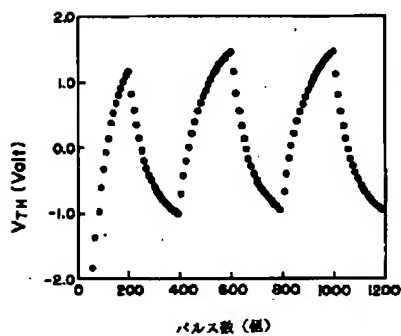


【図 2】

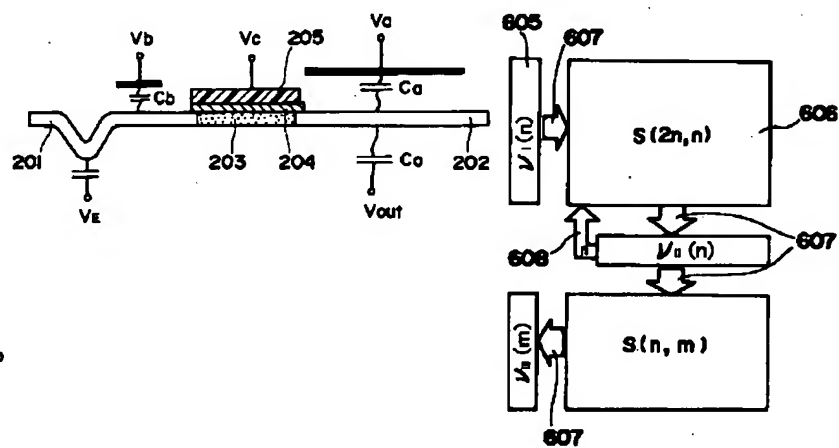


【図 6 b】

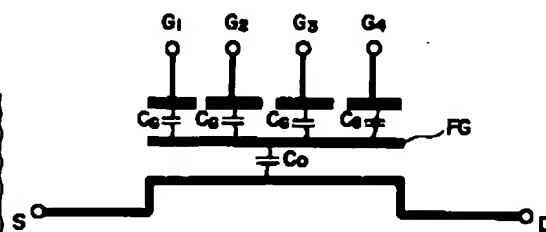
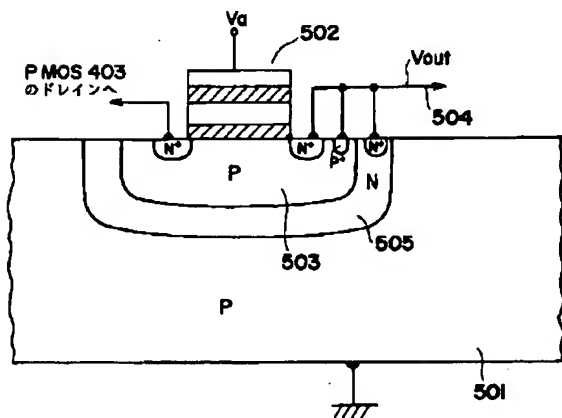
【図 1 j】



【図 5】

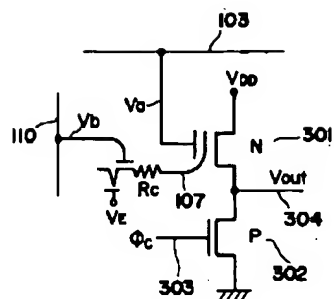


【図 11 b】

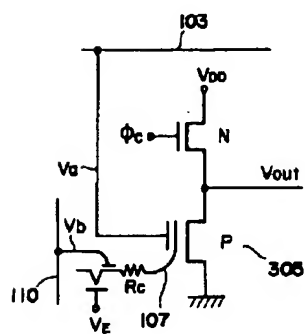


【図 3】

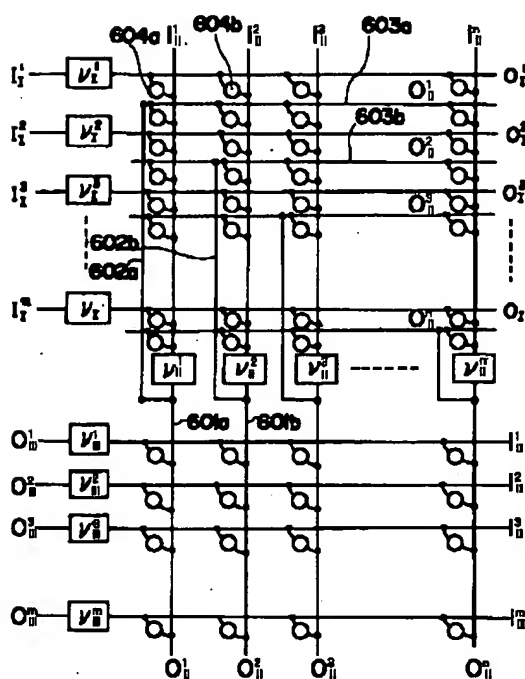
(a)



(b)

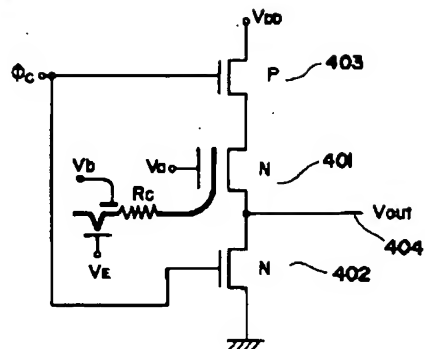


【図 6 a】

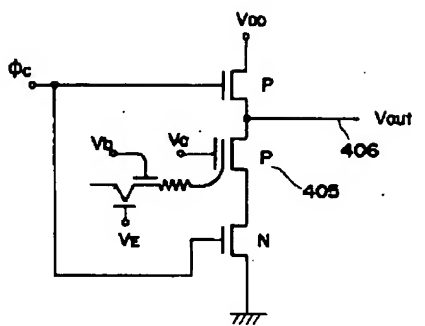


【図 4】

(a)

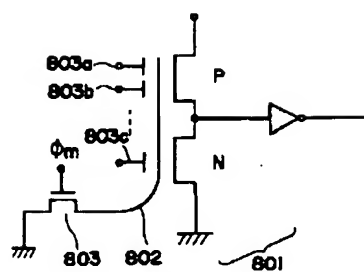


(b)

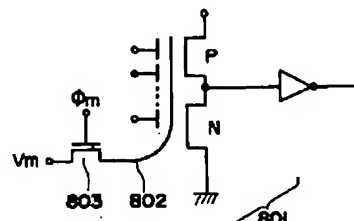


【図 8】

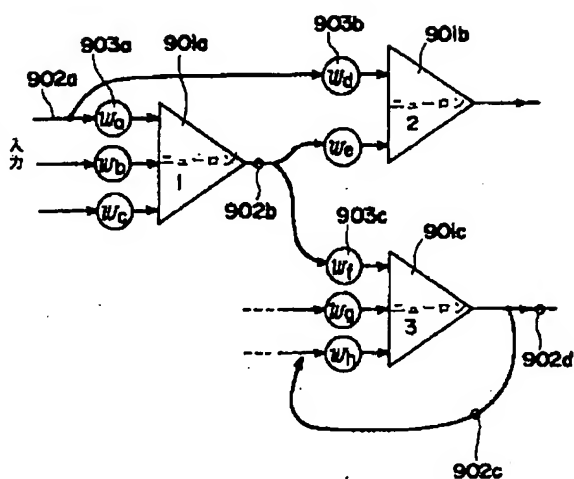
(a)



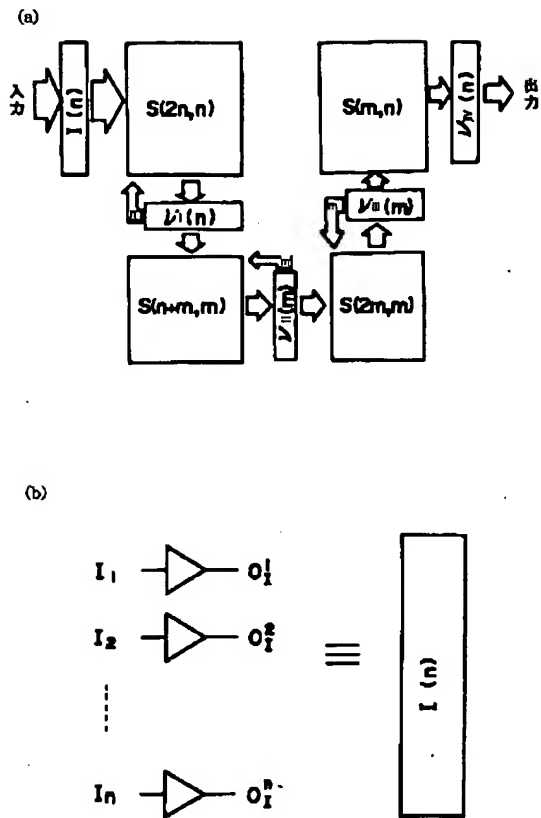
(b)



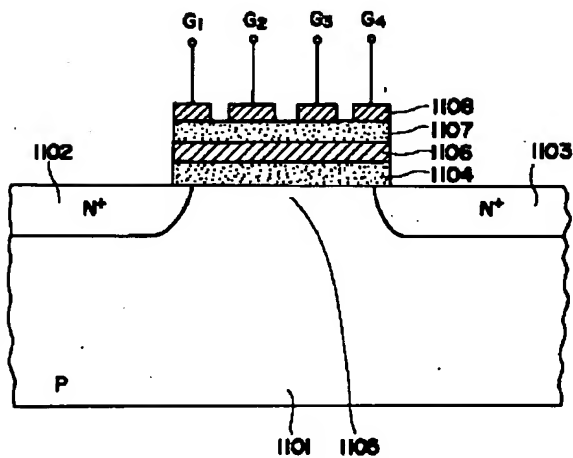
【図 9】



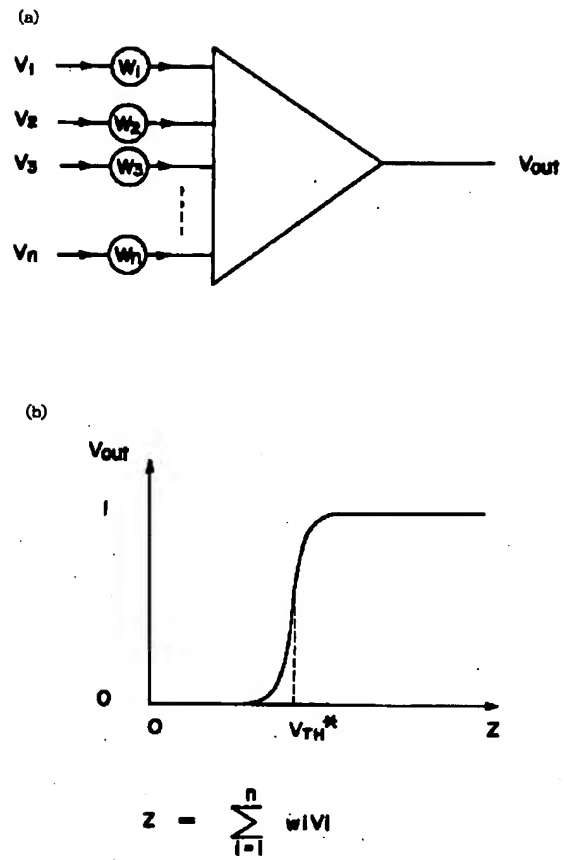
【図 7】



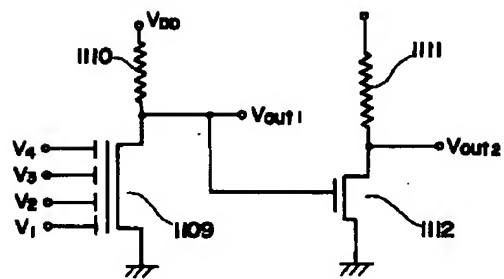
【図 11 a】



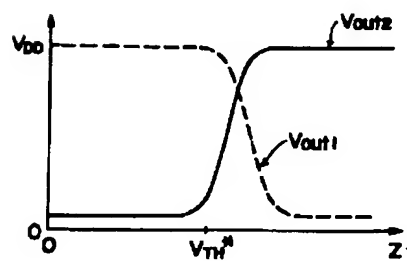
【図 10】



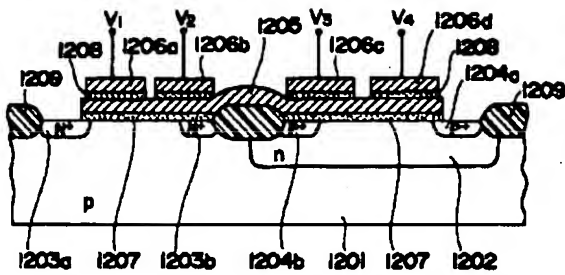
【図 11 c】



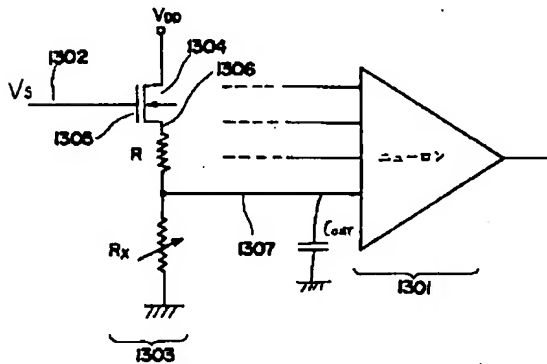
【図 11 d】



【図12a】

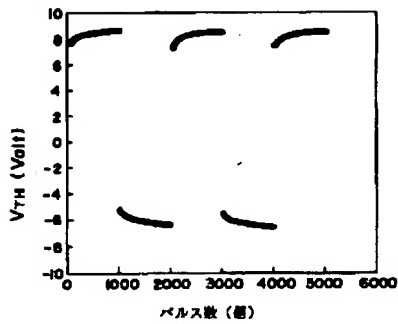


【図13】

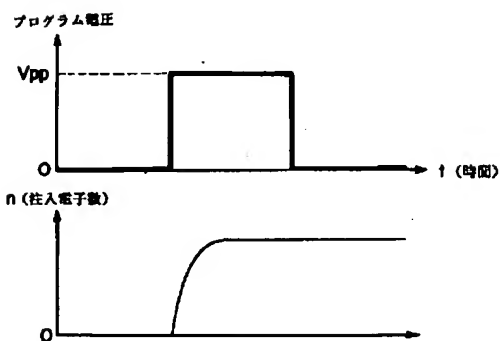


【図15】

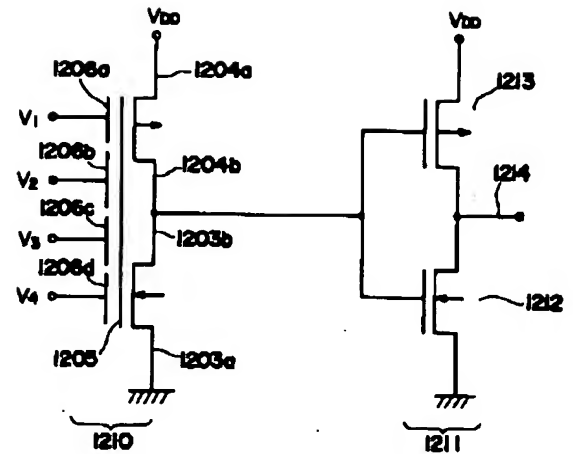
(a)



(b)

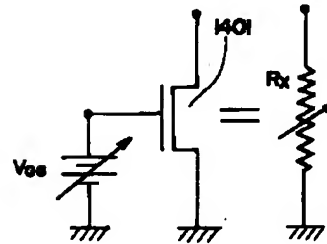


【図12b】

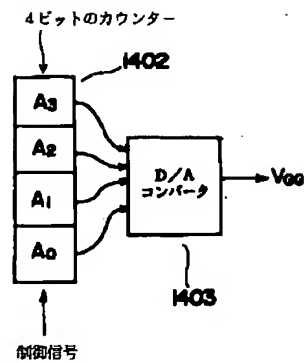


【図14】

(a)



(b)



フロントページの続き

(72)発明者 大見 忠弘
宮城県仙台市青葉区米ヶ袋 2-1-17-
301
(72)発明者 藤田 修
東京都千代田区内幸町 1 丁目 1 番 6 号日本
電信電話株式会社内

(72)発明者 森江 隆
東京都千代田区内幸町 1 丁目 1 番 6 号日本
電信電話株式会社内
(72)発明者 雨宮 好仁
東京都千代田区内幸町 1 丁目 1 番 6 号日本
電信電話株式会社内